

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 January 2001 (11.01.2001)

PCT

(10) International Publication Number
WO 01/02609 A2

- (51) International Patent Classification⁷: C12Q 1/68
- (21) International Application Number: PCT/US00/40292
- (22) International Filing Date: 30 June 2000 (30.06.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/142,335 2 July 1999 (02.07.1999) US
09/607,539 29 June 2000 (29.06.2000) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/142,335 (CIP)
Filed on 2 July 1999 (02.07.1999)
US 09/607,539 (CIP)
Filed on 29 June 2000 (29.06.2000)
- (71) Applicant (for all designated States except US): CURAGEN CORPORATION [US/US]; 11th floor, 555 Long Wharf Drive, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): GOULD-ROTHBERG, Bonnie, E. [CA/US]; 1701 Moose Hill Road, Guilford, CT 06437 (US). DIPIPO, Vincent, A. [US/US]; 156 North Atwater Street, East Haven, CT 06424 (US).
- 06512 (US). MILROY, Laura [US/US]; Building 3 Apt. B3, 175 Brushy Plain Road, Branford, CT 06405 (US). DANIELS, Kellye, K. [US/US]; 15 Manorwood Drive, Branford, CT 06405 (US).
- (74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— Without international search report and to be republished upon receipt of that report.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/02609 A2

(54) Title: METHOD OF IDENTIFYING TOXIC AGENTS USING DIFFERENTIAL GENE EXPRESSION

(57) Abstract: Disclosed are methods of identifying toxic agents, e.g., hepatotoxic agents, using differential gene expression. Also disclosed are novel nucleic acid sequences whose expression is differentially regulated by acetaminophen.

091917, 800 #13

METHOD OF IDENTIFYING TOXIC AGENTS USING DIFFERENTIAL GENE EXPRESSION

FIELD OF THE INVENTION

The invention relates generally to nucleic acids and polypeptides and in particular to
5 the identification of toxic agents in liver tissue using differential gene expression.

BACKGROUND OF THE INVENTION

Non-steroidal anti-inflammatory drugs (NSAIDs) are commonly used for controlling pain and fever in both adults and children. One type of NSAID agent is acetaminophen. Acetaminophen is well tolerated by most individuals and is reported to not induce side effects
10 sometimes observed with other commonly used pain relievers.

When high doses of acetaminophen are ingested, however, complications can result. One type of complication is liver damage, *e.g.* pericentral hepatic necrosis. Liver toxicity associated with acetaminophen ingestion is thought to occur when this agent is metabolized in the liver to N-acetyl-parabenzoquinoneimine, which can be toxic. In most healthy individuals,
15 this potentially toxic compound is degraded in the liver by the polypeptide glutathione. However, when acetaminophen is ingested at high doses, insufficient amount of glutathione may be present to process the N-acetyl-parabenzoquinoneimine. As a result, liver damage can result. In severe cases, potentially fatal hepatic necrosis may follow.

Acetaminophen poisoning can be treated by administering an antidote. The antidote is
20 preferably administered as soon as possible after poisoning. However, the antidote may not be promptly administered because signs of acetaminophen poisoning can be confused with flu-like symptoms. Early indications of acetaminophen poisoning can include, *e.g.*, nausea, and loss of appetite, diarrhea, abdominal pain, and vomiting. These flu-like symptoms are usually present between 4 to 12 hours after taking the drug. Later indications of acetaminophen
25 poisoning can include, *e.g.*, confusion, jaundice, or unconsciousness. However, by the time administering an antidote will have little effect.

SUMMARY OF THE INVENTION

The invention is based in part on the discovery that certain nucleic acids are differentially expressed in liver tissue of animals treated with acetaminophen. These differentially expressed nucleic acids include novel sequences and nucleic acids sequences that, while previously described, have not heretofore been identified as acetaminophen responsive. These differentially expressed nucleic acids can be used to identify agents that damage livers, *i.e.*, are hepatotoxic, *e.g.* pericentral hepatic necrosis. These nucleic acids can in addition be used to identify poisoning associated with ingestion of NSAIDS, *e.g.*, acetaminophen, in a subject.

10 In various aspects, the invention includes methods of method of screening a test agent for toxicity, *e.g.*, hepatotoxicity. For example, in one aspect, the invention provides a method of identifying a hepatotoxic agent by providing a test cell population comprising a cell capable of expressing one or more nucleic acids sequences responsive to acetaminophen, contacting the test cell population with the test agent and comparing the expression of the nucleic acids sequences in the test cell population to the expression of the nucleic acids sequences in a reference cell population. An alteration in expression of the nucleic acids sequences in the test cell population compared to the expression of the gene in the reference cell population indicates that the test agent is hepatotoxic.

20 In a further aspect, the invention provides a method of assessing the hepatotoxicity of a test agent in a subject. The method includes providing from the subject a cell population comprising a cell capable of expressing one or more acetaminophen-responsive genes, and comparing the expression of the nucleic acids sequences to the expression of the nucleic acids sequences in a reference cell population that includes cells from a subject whose exposure status to a hepatotoxic agent is known. An alteration in expression of the in the test cell population compared to the expression of the nucleic acids sequences in the reference cell population indicates the hepatotoxicity of the test agent in the subject.

25 In another aspect, the invention provides a method of diagnosing or determining susceptibility to hepatotoxicity. The method includes providing from the subject a cell population comprising a cell capable of expressing one or more ACETA responsive genes, and comparing the expression of the nucleic acids sequences to the expression of the nucleic acids

sequences in a reference cell population that includes cells from a subject not suffering from hepatotoxicity. An alteration in expression of the in the test cell population compared to the expression of the nucleic acids sequences in the reference cell population indicates subject has or is susceptible to a hepatotoxicity.

- 5 Also provided are novel nucleic acids, as well as their encoded polypeptides, whose expression is responsive to the effects of acetaminophen.

 Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can
10 be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

- 15 Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

 The present invention is based in part on the discovery of changes in expression patterns of multiple nucleic acid sequences in rat liver cells following exposure to
20 acetaminophen.

 The differentially expressed nucleic acids were identified by administering acetaminophen orally to male 10-14 week old Sprague Dawley rats at either an ED100 dose of 43.3 mg/kg/day for 72 hours, or an LD10 dose of 133.3 mg/kg/day for 24 hours. Differential gene expression was evaluated in two studies. In the first study, the control animals received
25 canola oil. In the second study, the control animals received 10% ethanol. After sacrificing animals at the designated time points, liver tissue was dissected, total RNA was recovered from the dissected tissue, and cDNA prepared.

 Sequences expressed in different levels in acetaminophen-treated and control rats were identified by subjecting the cDNA to GENE CALLING™ differential expression analysis as

described in U. S. Patent No. 5,871,697 and in Shimkets et al., Nature Biotechnology 17:798-803 (1999). The contents of these patents and publications are incorporated herein by reference in their entirety.

5 In the first study, over 400 gene fragments were initially found to be differentially expressed in rat liver tissue in response to administering acetaminophen. Genes fragments whose expression levels were modulated greater than 8-fold were selected for further analysis.

Differential expression of the gene fragments was confirmed using a unlabeled oligonucleotide competition assay as described in Shimkets et al., Nature Biotechnology 17:198-803. Forty-two confirmed sequences whose expression is modulated at least 8-fold in
10 the presence of acetaminophen were identified. These 42 nucleic acid sequences are referred to herein as ACETA 1-42, and are summarized in Table 1. The row subheading in Table 1 (e.g., protein production, signal transduction, etc.) provides a functional classification for the protein.

Ten sequences identified in the first study (ACETA: 1-10) represent novel genes.
15 Eight of these ten genes show some similarity to previously described sequences. Of these eight genes, three have $\geq 95\%$ identity, four genes have between 90%- 94% identity, and one gene has 81% identity, to known genes. The remaining novel genes, ACETA 6 and 7, do not appear to be related to previously described sequences. The remaining sequences (ACETA:11-42) have been previously described.

20 In the second study, 133 confirmed sequences were identified whose expression is modulated in response to acetaminophen. These sequences are summarized in Table 2. Six of the 133 nucleic acid sequences identified were also identified in the first study. These sequences are ACETA 15, 20, 23, 33, 36, 41. The remaining newly identified 127 nucleic acid sequences are referred to herein as ACETA 43-170.

25 Eight sequences identified in the second study represent novel genes (ACETA43-45, 48-51 and 54). Four sequences (ACETA 46, 47, 52 and 53) represent novel ESTs. Seven of the eight genes show some similarity to previously described sequences. Of these seven genes, two have $\geq 90\%$ identity, two genes have between 80%- 89% identity, and three genes have between 70%- 79% identity to known genes. One gene, ACETA 43 does not appear to be

related to previously described sequences. The remaining sequences (ACETA:55-170) have been previously described.

5 Newly described sequences are presented herein. For some of the novel sequences (i.e., ACETA: 1-10 and ACTETA 43-54), a cloned sequence is provided along with one or more additional sequence fragments (*e.g.*, ESTs or contigs) which contain sequences substantially identical to, the cloned sequence. Also provided for some sequences is a consensus sequence that includes a composite sequence assembled from the cloned fragments and additional fragments. For a given ACETA sequence, its expression can be measured using any of the associated nucleic acid sequences in the methods described herein.

10 For previously described sequences (ACETA:11-42 and ACETA 55-170) database accession numbers are provided. This information allows for one of ordinary skill in the art to deduce information necessary for detecting and measuring expression of the ACETA nucleic acid sequences. For example, primers for PCR amplification can be made based on the sequences corresponding to the database accession numbers, and/or on the sequences disclosed
15 herein.

The acetaminophen -responsive nucleic acids discussed herein include the following:

Table 1

<u>Description of Sequence</u>	Acetaminophen Effect on Transcript Level				
	<u>GenBank Acc#</u>	<u>LD10</u>	<u>ED100</u>	<u>ACETA Assignment</u>	<u>SEQ ID NO</u>
NOVEL GENES					
Novel gene fragment, 370 bp, 81 % S.I. To rat 5S rRNA gene	X83749	N.D.	-35x	1	1
Novel gene fragment, 81 bp, 98 % S.I. To rat genes for 18S, 5.8S, and 28S ribosomal RNAs	V01270	+180x	+65x	2	2 and 3
Novel gene fragment, 411 bp, 95% S.I. to rat alpha-tubulin	V01227	N.D.	-10x	3	4
Novel gene fragment, 108 bp, 90% S.I. to Human complement component C1r	X04701	+4x	+3.5x	4	5
Novel gene fragment, 348 bp, 93% S.I. to lambda phage lacZ translational fusion vector	U39284	N.D.	+50x	5	6
Rat novel gene fragment, 622 bp		-2x	-5x	6	7
Rat novel gene fragment, 143 bp		N.D.	-30x	7	8 and 9
Novel gene fragment, 308 bp, 94% S.I. to rat polymorphic marker D17UIA4 Sequence	AF054078	-15x	-30x	8	10 and 11
Novel gene fragment, 290 bp, 90% S.I. to mouse surfeit locus protein 4	Q64310	+2x	+5x	9	12
Novel gene fragment, 166 bp, 98% S.I. to human alpha-N- acetyl/galactosaminidase	M38083	-100x	-100x	10	13

<u>Description of Sequence</u>	<u>GenBank Acc#</u>	<u>LD10</u>	<u>ED100</u>	<u>ACETA Assignment</u>	<u>SEQ ID NO</u>
1.:PROTEIN PRODUCTION					
1.2.:mRNA TRANSLATION					
1.2.1.:RIBOSOMAL PROTEINS					
L18 protein	M20156	-5x	-5x	11	
Ribosomal protein L24	X78443	-5x	-5x	12	
1.2.6.:RIBOSOMAL RNAs					
18S rRNA gene possible contaminants	M11188	+300x	+50x	13	
Genes for 18S, 5.8S, and 28S ribosomal RNAs. possible contaminants	V01270	-50x	-10x	14	
1.5.:PROTEIN DEGRADATION					
1.5.3.:PROTEASOME COMPONENTS					
Proteasome RN3 subunit mRNA, complete cds.	L17127	N.D.	-10x	15	
2.:SIGNAL TRANSDUCTION					
2.1.:PEPTIDE HORMONES/GROWTH FACTORS/CYTOKINES					
2.1.5.:PEPTIDE HORMONE BINDING PROTEINS					
Fetuin	X63446	-35x	-35x	16	
2.14.:DNA BINDING PROTEIN					
2.14.2.:NUCLEAR HORMONE RECEPTORS					
Farnesoid X activated receptor	U18374	+80x	+5x	17	
Small heterodimer partner homolog	D86580	N.D.	-40x	18	
3.:CELL CYCLE					
3.6.:CELL DEATH REGULATION					
3.6.2.:APOPTOSIS INHIBITION					
Testis enhanced gene transcript	X75855	+70x	+85x	19	
4.:BASIC METABOLISM					
4.1.:LIPID METABOLISM					
4.1.1.:FATTY ACID SYNTHESIS					
Stearyl-CoA desaturase	J02585	+3x	+2x	20	

<u>Description of Sequence</u>	<u>GenBank Acc#</u>	<u>LD10</u>	<u>ED100</u>	<u>ACETA Assignment</u>	<u>SEQ ID NO</u>
4.1.3.:KETONE BODY METABOLISM					
3-hydroxy-3-methylglutaryl-CoA synthase	M33648	-2x	N.D.	21	
4.2.:STEROID METABOLISM					
4.2.1:CHOLESTEROL BIOSYNTHESIS					
Farnesyl pyrophosphate synthetase	M34477	-2x	-3x	22	
4.2.4.:EXCRETION					
Cytochrome P-450(M-1)	J02657	O	-10x	23	
Estrogen sulfotransferase (Ste1) mRNA, complete	U50204	+200x	+100x	24	
4.3.:CARBOHYDRATE METABOLISM					
4.3.1.:GLYCOLYSIS/GLUCONEOGENESIS					
Aldolase B	M10149	-2x	N.D.	25	
Phosphoenolpyruvate carboxykinase (GTP) gene	K03248	-70x	N.D.	26	
4.3.2.:GLYCOGEN MANIPULATION					
Glycogen phosphorylase muscle isozyme	L10669	-50x	-40x	27	
4.4.:OXIDATIVE PHOSPHORYLATION					
4.4.1.:CITRIC ACID CYCLE					
Succinyl-CoA synthetase alpha subunit	J03621	-2x	-2x	28	
4.4.2.:ELECTRON TRANSPORT CHAIN					
Cytochrome oxidase subunits I,II, III genes	J01435	-40x	N.D.	29	
4.5.:AROMATIC AMINO ACID FAMILY					
4.5.7.:UREA CYCLE					
Arginase	J02720	-2x	-4x	30	
4.5.8.:METHYL CYCLE					
S-adenosyl-L-homocysteine hydrolase	M15185	-3x	-3x	31	

<u>Description of Sequence</u>	<u>GenBank Acc#</u>	<u>LD10</u>	<u>ED100</u>	<u>ACETA Assignment</u>	<u>SEQ ID NO</u>
4.7.: BIOSYNTHESIS OF COFACTORS, PROSTHETIC GROUPS, CARRIERS					
4.7.3.: HEME & PORPHYRIN					
Delta-aminolevulinic acid dehydrogenase	X04959	N.D.	-2x	32	
4.9.: DETOXIFICATION					
4.9.3.: HEAVY METALS					
Metallothionein-I (mt-1)	J00750	-3x	-2x	33	
4.9.5.: HYDROCARBONS					
Glutathione S-transferase Ya subunit	K01931	N.D.	+1.8x	34	
4.11.: METABOLITE STORAGE/TRANSPORT PROTEINS					
4.11.1.: EXTRACELLULAR TRANSPORT					
4.11.1.2.: LIPIDS					
Apolipoprotein C-III	J02596	-30x	-24x	35	
4.11.1.3.: STERIODS					
Transferrin	X77158	+360x	+380x	36	
4.11.1.4.: HYDROCARBONS					
Alpha-2-u globulin	M27434	-30x	-60x	37	
5.: TISSUE ARCHITECTURE					
5.1.: CYTOSKELETON					
5.1.1.: COMPONENTS					
5.1.1.1.: MOTOR ARM					
5.2.: EXTRACELLULAR MATRIX					
5.2.1.: COMPONENT					
Collagen alpha 1 type III	AJ005395	O	-15x	38	
5.2.3.: BREAKDOWN INHIBITION					
Alpha-1-antitrypsin mRNA, complete cds.	M32247	+50x	N.D.	39	
Contrapsin-like protease inhibitor (CPI-21)	D00751	-15x	-15x	40	
6.: EXTRACELLULAR ENVIRONMENTAL REGULATION					
6.2.: IMMUNE SYSTEM					
6.2.1.: COMPLEMENT					
6.2.1.1.: COMPONENTS					
pcRC201 mRNA for pre-pro-complement C3	X52477	+400x	+550x	41	

<u>Description of Sequence</u>	<u>GenBank</u> <u>Acc#</u>	<u>LD10</u>	<u>ED100</u>	<u>ACETA</u> <u>Assignment</u>	<u>SEQ ID NO</u>
7.: CELLULAR ORGANELLE STRUCTURAL INTEGRITY					
7.2.:ENDOPLASMIC RETICULUM/GOLGI APPARATUS					
7.2.2.:TRANSMEMBRANE PROTEINS					
7.4.: PEROXISOME/LYSOSOME/ENDOSOME					

9.:UNKNOWN FUNCTION

9.1.:KNOWN GENES

9.1.2.:UNASSOCIATED

L1 retrotransposon ORF2	U83119	O	-4x	42	
-------------------------	--------	---	-----	----	--

N.D. = Not Determined

TABLE 2.

<u>Gene Discovered</u>	<u>GenBank Acc. #</u>	<u>Acetaminophen Effect on Transcript Level</u>	<u>ACETA Assignment</u>	<u>SEQ ID NO</u>
Novel Genes				
Novel gene fragment, 228 bp	N/A	-11.7	43	14
Novel gene fragment, 571bp. 90% SI to Human beta-tubulin [X79535]	N/A	2	44	15 and 16
Novel gene fragment, 619bp, 79% SI to guanylyl cyclase A / atrial natriuretic peptide receptor [J05677]	N/A	-3.5	45	17 and 18
EST193210 Normalized rat ovary	AA850443	-1.6	46	19
EST223737 Normalized rat spleen	AI180006	-1.5	47	20
Novel gene fragment, 138bp. 89% SI to Mouse complement factor H-related protein [M29009]	N/A	4.3	48	21
Novel gene fragment, 419bp, 76% SI to mouse voltage-gated sodium channel [AF121804]	N/A	-2.9	49	22
Novel gene Fragment, 109 bp 89% SI to Bacteria binding macrophage receptor MARCO		-1.6	50	23
Novel gene fragment, 299bp, 73% SI to Chinese hamster glutathione S-transferase subunit [L20466]	N/A	-1.6	51	24
EST197321 Normalized rat liver	AA893518	4.4	52	25
EST227997 Normalized rat embryo	AI231309	3	53	26
Novel gene fragment, 205bp, 90% SI to mouse C1 inhibitor gene [AF052039]	N/A	2.2	54	27
Previously Described Genes				
17-beta hydroxysteroid dehydrogenase type 2	X91234	-4.5	55	
3 alpha-Hydroxysteroid dehydrogenase	D17310	2	56	
60 kDa protein and non-specific lipid transfer protein	M62763_0	4.3	57	
Adenylate kinase 3	D13062	-2.5	58	
Alanine aminotransferase	D10354	-2.8	59	
Aldehyde dehydrogenase	M73714	1.6	60	
Aldehyde oxidase male form (AOX1)	AF110478	1.9	61	
Alpha albumin	X76456	2.3	62	
Alpha-1-macroglobulin	M77183	4.4	63	
Alpha-2u-globulin (S type)	M26837	-13.3	64	
Alpha-fibrinogen.	X86561	1.8	65	
Alpha-tocopherol transfer protein	D16339	4.9	66	
Apolipoprotein B	M27440	8	67	
Argininosuccinate lyase	D28501	-2.4	68	
Beta-2 glycoprotein I	X15551	-3.7	69	
Beta-alanine oxoglutarate	D87839	1.5	70	
Beta-galactoside alpha 2,6-sialyltransferase	M18769	3.7	71	
Betaine-homocysteine methyltransferase (BHMT)	U96133	2.8	72	
Bile canaliculus domain-specific membrane glycoprotein	J02997	8	73	
C4BP alpha chain protein	Z50051	3.2	74	
Calnexin	L18889	3	75	
Calreticulin	X79327	2.7	76	
Canalicular multidrug resistance protein	X96393	-1.8	77	
Carboxyl methyltransferase	M26686	2.6	78	

<u>Gene Discovered</u>	<u>GenBank</u> <u>Acc. #</u>	<u>Acetaminophen</u> <u>Effect on</u> <u>Transcript Level</u>	<u>ACETA</u> <u>Assignment</u>	<u>SEQ ID</u> <u>NO</u>
Carboxylesterase	M20629	-1.6	79	
Cathepsin B	X82396	-2.3	80	
CDK109	Y17323	4.5	81	
Ceruloplasmin	L33869	8	82	
Chymotrypsin B (chyB)	K02298	-6.3	83	
Connexin protein Cx26	X51615	2.4	84	
C-reactive protein	M83176	4.8	85	
CYP2A2 (Testosterone 15-alpha-hydroxylase)	M33313 / J04187	2	86	
CYP2B2 (Cytochrome p450e phenobarbital inducible)	J00720	4	87	
CYP2B2 (Cytochrome p450e phenobarbital inducible)	K00996 / M37134	5	88	
CYP2C11 (Cytochrome p450h)	J02657	-1.6	23	
CYP2C13 (Cytochrome p450g)	M32277	1.9	89	
CYP2C23 (Arachidonic acid epoxigenase)	U04733	2.2	90	
CYP2C23 (Cytochrome p450 c117)	X55446	2.2	91	
CYP2C6	K03501	-1.5	92	
CYP2D5 (Cytochrome p450 isozyme CMF1b)	M22329	3.3	93	
CYP2E1 (Diabetes-inducible)	S48325 / M20131	-3.9	94	
CYP3A18 (Cytochrome p450/6 beta-2)	X79991	5	95	
CYP4A1 (Cytochrome p452)	X07259/ M14972	3.3	96	
CYP4A2	M57719	2.5	97	
CYP4F1 (Hepatic tumor cytochrome p450)	M94548	2	98	
CYP8B (Sterol 12alpha-hydroxylase P450)	AB009686	-10	99	
Cystathionine gamma-lyase	D17370	3	100	
Cytosolic 3-hydroxy 3-methylglutaryl coenzyme A	X52625	3	101	
Cytosolic malate dehydrogenase (Mdh)	AF093773	2.3	102	
Cytosolic NADP-dependent isocitrate dehydrogenase	L35317	3.5	103	
D-binding protein	J03179	-10	104	
Delta 4-3-ketosteroid-5-beta-reductase	D17309	3	105	
Dihydropyrimidinase	D63704	1.5	106	
Elongation factor 2	Y07504	2.2	107	
Estrogen sulfotransferase (Ste2)	U50205	2	108	
Eukaryotic initiation factor 5 (eIF-5)	L11651	2.9	109	
GADD45gamma	AB020978	1.6	110	
Gene 33 polypeptide	X07266	5	111	
Glucose transporter type 2	L28135	-1.8	112	
Glu-Pro Dipeptide Repeat	U40628	-3	113	
Glutathione S-transferase Ya subunit	M26874	1.5	114	
Glutathione S-transferase Yc2 subunit	S72506	1.6	115	
Glyceraldehyde-3-phosphate-dehydrogenase (GAPDH)	X02231	1.6	116	
Glycogen synthase	J05446	-2.5	117	
Grp75=75 kda glucose regulated protein	S78556	3.2	118	
Haptoglobin	K01933	-4.3	119	
Hydroxysteroid sulfotransferase	M31363	1.6	120	

<u>Gene Discovered</u>	<u>GenBank Acc. #</u>	<u>Acetaminophen Effect on Transcript Level</u>	<u>ACETA Assignment</u>	<u>SEQ ID NO</u>
Interleukin 6 receptor ligand binding chain	M58587	-3.4	121	
Lanosterol 14-alpha-demethylase	U17697	2.8	122	
Leuserpin-2.	X74549	2.4	123	
L-gulono-gamma-lactone oxidase	J03536	1.6	124	
Lipopolysaccharide binding protein	L32132	-2	125	
Lysosomal acid lipase, intracellular hydrolase	S81497	2.2	126	
Medium chain acyl-CoA dehydrogenase	J02791	1.8	127	
Metallothionein-i (mt-1)	J00750	-6.6	33	
MHC-associated invariant chain gamma	X13044	-1.7	128	
Mitochondrial acetoacetyl-CoA thiolase	D13921	3.5	129	
Mitochondrial aldehyde dehydrogenase	X14977	2.8	130	
Mitochondrial enoyl-CoA hydratase	X15958	-1.6	131	
Mitochondrial succinyl-CoA synthetase alpha subunit	J03621	2.4	132	
NAD+-isocitrate dehydrogenase, gamma subunit	X74125	-2.6	133	
NADPH-cytochrome P-450 reductase	M12516	-2.2	134	
Nucleus-encoded mitochondrial carbamyl phosphate synthetase I	M12335	3.5	135	
Ornithine aminotransferase	M11842	1.5	136	
Orphan nuclear receptor OR-1	U20389	-2	137	
P450 6beta-2	D38381	5	138	
PcRC201 pre-pro-complement C3	X52477	3.9	41	
Peroxisomal enoyl-CoA: hydratase-3-hydroxyacyl-CoA bifunctional enzyme	K03249	2.5	139	
Phospholipid hydroperoxide glutathione	X82679	-1.6	140	
Phosphoribosylpyrophosphate	D26073	-2.3	141	
Phosphoribosylpyrophosphate	D26073	-2.3	142	
Phosphotidylethanolamine N-methyltransferase	L14441	-1.6	143	
PKC-zeta-interacting protein	Y08355_1	-1.8	144	
Plasma protein	Y11283	4.2	145	
Plasma proteinase inhibitor alpha-1-inhibitor III	J03552	5	146	
Plasma proteinase inhibitor alpha-1-inhibitor III group 3	M22360	5	147	
Polypeptide 1-microglobulin/bikunin	S87544	-3	148	
Prealbumin (transthyretin)	K03251	4.3	149	
Preproalbumin	V01222	20	150	
Pre-pro-complement C3	X52477	-9	151	
Pro alpha 1 collagen type III	X70369	4	152	
Proteasome RN3	L17127	-1.6	15	
Protein kinase inhibitor p58		2.1	153	
Putative glycogen storage disease type 1b protein	AF080468	-1.7	154	
rat 8s ma	K01592	-4.4	155	
Rat ortholog of M. musculus mRNA Poly(A) binding protein, 259 bp [X65553]		2.4	156	
Rat ortholog of Mus musculus mRNA for Alix (ALG-2-interacting protein X), 248bp [AJ005073]		-1.7	157	
Ribophorin I	X05300	2.8	158	
Serine protease	D88250	1.6	159	
Small zinc finger-like protein (TIM13)	AF144701	-2	160	
Stearyl-CoA desaturase	J02585	10	20	

<u>Gene Discovered</u>	<u>GenBank</u> <u>Acc. #</u>	<u>Acetaminophen</u> <u>Effect on</u>	<u>ACETA</u>	<u>SEQ ID</u>
		<u>Transcript Level</u>	<u>Assignment</u>	<u>NO</u>
Sterol 12-alpha hydroxylase	AB018596	-1.7	161	
Sulfated glycoprotein 2	X13231	1.5	162	
Testosterone 6-beta-hydroxylase, cytochrome	U09725	3.1	163	
Transferrin	X77158	1.5	36	
Trihydroxycoprostanoyl-CoA Oxidase	X95189	-1.7	164	
Tumor necrosis factor receptor (TNF receptor)	M63122	-1.5	165	
UDP-glucuronosyltransferase	M31109	5.5	166	
UDP-glucuronosyltransferase, phenobarbital-inducible form	M13506	4	167	
UI-R-AA0-wq-a-05-0-UI.s1 UI-R-AA0	A1578698	-2.6	168	
Urinary protein 2 precursor	AF198441	4.4	169	
Very-long-chain acyl-CoA synthetase	D85100	3.1	170	

Below follows additional discussion of the novel nucleic acid sequences whose expression is differentially regulated in the presence of acetaminophen.

5 ACETA1

ACETA1 is a novel 370 bp gene fragment. The nucleic acid has the following sequence:

```

10      1   GGATCCCAGA CACTGGATGG AATGTTAGAG TTTGATTTTG CTTTGATTG TGACCGTGCC
      61CTGATATTTT TTCCCTCTTG AAGGAAGAAA GTATTTTAGT GGAGCCCACA GTTAAGAGAC
      121 TTTGAATTGA AAAAAAGACT TTGAATTTTA AAAGAGATTG GATATTTAAT GGGATTGAAA
      181 TTTTAAATG TAAAGACTGT GGGACTCTTA AAGTTATTTA GACCTTGGGG ATGAATAATA
      241 AAGTAAGGGT TGAGGTTTAA TAGTGATGEG GTTGTGTGTC AAATTGACAA GGGGTCAATT
      301 GTACTGACTG GCTTTGNGTG TCAACTTAAC TCAANCTGGA GTTNTCNCAG AAAAAGGAGC
15      361 CTCAGCCCCC (SEQ ID NO:1)

```

ACETA2

ACETA2 is a novel 436 bp gene fragment. The nucleic acid was initially identified in a 81 bp cloned fragment having the following sequence:

```

20      1   TCCCGGGTTA AGGCGCCCGA TGCCGACGCT CATCATACCC CAAAAAGTG TTGGTTGATA
      61   TANACAGCAG GACGGGGGCC R (SEQ ID NO:2)

```

The cloned sequence was assembled into a contig resulting in the following consensus sequence:

```

25      1   TCGGGCCGCTCCCGTCCCGTCCGACTGCCGGCGACGGCCGGGTATGGGCCCGACGCTCCAGCGCCATCCATTTTCAGGG
      81   CTAGTTGATTCGGCAGGTGAGTTGTTACACACTCCTTAGCGGATTCCGACTTCCATNGGCCACCGTCTGCTGTCTATAT
      161  CAACCAACACCTTTTCTGGGGTATGATGAGCGTCGGCATCGGGCGCCTTAACCCGGCGATTTCGGTTCATCCCGCAGCGCC
      241  AGTTCTGCTTACCAAAAGTGGCCCACTAGGCACTCGCATTCACGCCCGGCTCCAGCCAGCAGCGCGGGCTTCTTACCC
      321  ATTTAAAGTTTGAGAATAGGTTGAGATCGTTTCGGCCCCAAGACCTCTAATCATTCGCTTTACCGGATAAACTGCGGGT
30      401  TGTCGAGCGACCGCGTTGCCAGAGCGCCAGCTATCCTG (SEQ ID NO:3)

```

ACETA3

ACETA3 is a novel 411 bp gene fragment. The nucleic acid has the following sequence:

1 NNCTCNCTTATTTTCCANTTTTTTCTTCCATNATTCCTCCCTCAGGGTTTGATGNNGCCCTGAATGTTGATCTGACAG
5 81 AATTCCAGACCAACCTGGTGCCCTACCCTCGAATCCACTTCCCTCTGGCCACTTATGCCCTGTCTCTGCTGAGAAA
161 GCCTACCATGAGCAACTTACAGTAGCAGAGATCACCAATGCCTGCTTTGAGNCAGCCAACCAGATGGTGAAATGTGACCC
241 TCGCCATGGTAAATACATGGCTTGCTGCCTGTACCGTGGTGATGTGGTCCCCAAAGATGTCAATGCTGCCATTGCCA
321 CCATCAAGACCAAGCGCACCATCCAGTTTGTGGACTGGTGCCCACTGGCTTCAAGGTTGGCATTAAATTACCAGCCTCCC
401 ACTGTGGTACC (SEQ ID NO:4)

ACETA4

ACETA4 is a novel 108 bp gene fragment. The nucleic acid has the following sequence:

15 1 GTGCACCTCA AGTTCCTGGA TCCTTTTGAA ATTGATGACC ACCAGCAAGT ACACTGCCCC
61 TATGACCAGC TCCAGATCTA CGCTAATGGG AAGAACTGG GTGAATTC (SEQ ID NO:5)

ACETA5

20 ACETA5 is a novel 348 bp gene fragment. The nucleic acid has the following sequence:

25 1 TGTACACTGC AGCCTCGGTA TCCAGCACAA CCTGCACGGA CAGGCCGGTA TATGCCGACA
61 CCTTCTGCGC AAACATCTGG CGGGTTGCGT CCATCCGGGA CTGCAGTGTC TCCCCGGACG
121 TCATTCCCGG GAAAAAATGG GCTTGTAGGG GTTTGGCCC ATTCCACCC TTTATTGGGC
181 TTGCCCCGCTT GTAAAAATTC AANCGGNGGA ATTTCCACAC ACCCCTNGTT TNTTCCAAGC
241 CGCCANCCAC CCGTAAATTT ACTGGGGANC CCATTCATNG AACGCCCCGAT GGAACCTNG
301 TTCCGGGGCC GGTNTNNGGT NAACCCNNAA NCCCCCCCCC CCCCCCCC (SEQ ID NO:6)

ACETA6

ACETA6 is a novel 622 bp gene fragment. The nucleic acid has the following sequence:

```

1      CCTCATCAGCGTGATGAAAGGGAAAGATGTCTTTACCCAAGTTCCTTGTCTCACGGAGACGGGAAGACTAACAAGAAGG
5      81      CTTCTTAACATCGTAAAGACATGAGTTCGGTTGAACATAAGTATATAGGTTGGGTTAGAATGAACGGGTCCGTGTGGGTTG
      161     GTCTAGTGAAGAGATGAGCTAGATGCAATGTGTCATTTAGCGTTATGTCTTTAACCGGTGGGCTGCTGTAAGAATCGGTG
      241     GCAGTTCTCTCTGCGGTGTGTTAATTGCTCTGGAACGCTACTAGGACCCGAATACTAAGGCCACATCTCTACGTCTCT
      321     AAGAAGGGGAAATAAGATAGGCTTTAGTCTCACAGTGTGGCCTAGGTGGGGTTCATGTTACTCCCTAACACGCTACAGA
      401     ATTCAGACATAGTTTCTGTGTGGTGGTAGGTCTGTGCTTTTTTATCGTCTTGCTGCCATCTTCCAGCCAGTGTATGGT
10     481     TTATCCAGTCTGTGTGCCAAGCCAGCCATGTCTCCACGACCCGTTAGTCCAGAGGAGTTCTGCCCCAACACTAGTTTC
      561     CAGCTGCCCGCTCCTAATGTACACCAATCAAGACAGAATAAAATTGAGTTGTTTCGGTGCA (SEQ ID NO:7)

```

ACETA7

ACETA7 is a novel 642 bp gene fragment. The nucleic acid was initially identified in a 143 bp cloned fragment having the following sequence:

```

1      GGGCCCGAAA AGCAAAGAAC CCCCTGATGC TCCCCGCTGA GACTCACTAG CAGGGTTCCA
61     CGGGGTACGG TCCCCTGCAG TAGATGGGAG GTGGNGGGCA TTGGGAAGGC ACAGACAATC
20    121     AAATGTAGAC CGGCTAATAA AGT (SEQ ID NO:8)

```

The cloned sequence was assembled into a contig resulting in the following consensus sequence:

```

25      1      NTGTACACTTCGCCCAGTTTCTAAAAGGAAGATTCAATCTATGTGCCCTGACTGCCCGCATCCCGTTGACTTGTACAGCCC
      81      CCAGTGTTCTGGAAGCTGCCACAGAATCACTTGCGAAGTTTAACAGCGAGAACCCCTCAAAACAATATGCACTCGTCAAA
      161     GTCACNCAAGGCTACGACCCAGTGGGTAGTTGGTCCTTCTTACTTTGTGGAATATTTGATCAAAGAGTCACCATGTACCC
      241     AGTCTCAGGACAGCTGTTCACTCCAGGCCTCCGACTCTGAGCCCGTTGGTCTTTGCCAAGGTTCACTGATTAAGTCCC
      321     GGGGTCCCTCCTCAACGCTTTAAAAAGACTGTCACTGTGTCGTGCGAGTTTTTCGAATCTCAGGACCAGGTCCCTGGAGG
30     401     TGAGAACCCTGCTGATACCAAGATGCTAAGAACTCCCTCAGAAAAACACAGCCCTACCAGCTCACCTCCATAACTG
      481     CACCAAGAGGATCTATCCAACACCTCCCTGAGCAGGAGGAGCCTGAAGACTCCAAGGGAAGAGTCTGAGGAACCCCTT
      561     CCTGTGACGTGGATCTAACCACAAACCCACAGGGTGACACACTGGATGTCTCCTTCTCTACCTGGAGCCTGAGGAAAA
      641     GAAACTGGTGGTCTGCCTTTCCCTGGGAAGGAACAGCGCTCCCTGAGTGCCCGGGGCCGAAAGCAAAGAACCCCTT
      721     GATGCTCCCCGCTGAGACTCACTAGCAGGGTTCACGGGGTACGGTCCCTGCAGTAGATGGGAGGTGGTGGGCATTGGG

```

801 AAGGCACAGACAATCAAATGTAGACCGGCTAATAAAGTGTGTCCTTTGGATGCTTCTTGATCTCAAAAAAAAAAACTGCA
 881 GATATACNATACATCTNCAGNCCCCNTT (SEQ ID NO:9)

ACETA8

- 5 ACETA8 is a novel 411 bp gene fragment. The nucleic acid was initially identified in
 a 308 bp cloned fragment having the following sequence:

1 TGATCACACA AGAAAATACA GGGTCTACAG ATATCTGTCT CATTCACTCT ATCTTACTTG
 61 TAGATATTTG GGGAGATTGT AGATNGATAG ATAGATNGAT AGATAGATAG ATAGATAGAT
 10 121 AGATCGATAA TAGATAGACG ATACCTAGAT GGATAGATAG ATGATAGATA GATNGATAGA
 181 TAGATAGATA GATAGATAGG ATAGGGAGAT GGAGACAGAG AAGGAGGTGA AGCTGGGGGT
 241 AGAGATAGAG ATAGACAGAG ATATCTNACA TCCTAAACGT GTCTGNTTNA TTCTCAATAT
 301 TCTGTACA (SEQ ID NO:10)

- 15 The cloned sequence was assembled into a contig resulting in the following consensus
 sequence:

1 TGTACAGAATATTGAGAATNAANCAGACACGTTTAGGATGTNAGATATCTCTGTCTATCTCTATCTTACCCCCAGCTTC
 81 ACCTCCTTCTCTGTCTCCATCTCCCTATCCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCCAT
 20 161 CTAGGTATCTGTCTATCTATTATCGATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTACAATCTCCCC
 241 AAATATCTAAAAGTAAGATAGAGTGAATGAGACAGATATCTGTAGACACTGTATTTTCTTGTGTGATCAGATCTAGTGTG
 321 GTGGATGATAGAAGTTGAACTTGCTTTATTGCTATGTGTTAAATATTTGTTTGCAATAAATGGCCTTTTGAAATGCT
 401 TTTCTGTTCCCT (SEQ ID NO:11)

25 **ACETA9**

- ACETA9 is a novel 290 bp gene fragment. The nucleic acid has the following
 sequence:

1 YGGCCGAGGA TTTCGCCGAC CAGTTCCTTC GAGTCACCAA GCAGTACCTG CCTCATGTGG
 30 61 CACGCCTCTG CCTGATTAGC ACCTTCCTGG AAGATGGCAT CCGCATGTGG TTCCAGNGGA
 121 GNGAGCAGCG CGACTATATT GATACCACCT GGAGCTGGGG TTACCTGCTG GCCTCATCCT
 181 TTTCGTGGTT NCCCTCAAAC CCCTGCTTGG GGGGAACAAG GNNGGGAAAC TGGGCTTGGA
 241 GTTTNTTGNA NGNTGAAACN GGGGAAANTT TNGGCGNCGG GGGGNGAGNN (SEQ ID NO:12)

ACETA10

ACETA10 is a novel 116 bp gene fragment. The nucleic acid has the following sequence:

5 1 TGATCACTGT GAAGTTGGTT TTATCTCGGA GGCCACTGAT GATGTCACCT GAGTAGACGT
 61 CCTGGGCCTC ATATATCACA GACCCAGTGA AGTTCAGCTG GCCAAGGGAG GAGTGGTAGC
 121 GATAAGGCAT ATCGGTCCTG CAGCTGAAGA AGACTAAGGC GCTAGC (SEQ ID NO:13)

ACETA43

10 ACETA43 is a novel 391 bp gene fragment. The nucleic acid has the following sequence:

 1 TCTAGAAGTCCTTGGGCCCCAGGGGTCTCTGGGAAGTAGCAGCGACACCGTGGAACACACACAGCTTGTGATGGAGC
 81 TGTGGGTACCTCACAGCCCTCAGCCAGAGAAACAGTCTCTCTTGCCTGTTGTCTCCTCGCTGCCACATCGTCTTTGA
 161 GTGACTCCGGAAGATCCCGTTACAGGTAAGATCCCAGGATTTCCAGAACATCTTATCTGCAGTTTCTGTGATTATGCCAG
 15 241 GAGACACAATTGAGTGCAGACCTTACCAGAACAAGGAGCCACCAAGCTACCATCTCAGGGATAAAATAACTCCCACCATT
 321 GACTCAAGTTGCTTCAGTTTCTGAAGAACCATAGACTTCCATTGTCTTCTAGGTTTCTAAGAAGAATTC (SEQ ID
 NO:14)

ACETA44

20 ACETA44 is a novel 571 bp gene fragment. The nucleic acid was initially identified in a 411 bp cloned fragment having the following sequence:

 1 GGATCCCCAACAAATGTGAAGACGGCCGTGTGTGACATCCCTCCTCGTGGCCTCAAGATGTCAGCCACCTTCATTGGCAAC
 81 AGCACCGCCATCCAGGAGCTGTTCAAGCGCATCTCGGAGCAGTTCACTGCCATGTTTCAGGCGCAAGGCTTTCCTGCACTG
 161 GTACACGGGCGAGGGCATGGACGAGATGGAGTTCACCGAGGCAGAGAGCAACATGAATGACCTGGTGTCTGAGTACCAGC
 25 241 AGTACCAGGATGCCACGGCTGATGAGCAGGGCGAGTTCGAGGAGGAGGAGGCGAGGATGAGGCTTAAGAACTTCTCGGA
 321 TACATTGTGCACCCTTAGTGAACCTTCTGTTGTCTCCAGCATGGTCTTTCTATTGTAAATTATGGTGCTCAGTTTGCCT
 401 CTGTCTGAAAT (SEQ ID NO:15)

30 The cloned sequence was assembled into a contig resulting in the following consensus sequence:

 1 TTTTTTTTTTTTTCTTAAACACATGCTTTTTTATTCATATAGATTTCTGAGACAATACTGTAATCTTGAAAGGAGGTTTC
 81 ACACTATTACATCAACAGTGAATTTACAGACAGAGGCAAACTGAGCACCATAATTTACAAATAGAAAGACCATGCTGGAGG
 161 ACAACAGAAGTTCACCTAAGGTGCACAATGTATCCGAGAAGTTCTTAAGCCTCATCCTCGCCCTCCTCCTCGAACTC

241 GCCCTGCTCATCAGCCGTGGCATCCTGGTACTGCTGGTACTCAGACACCAGGTCATTGTTGCTCTCTGCTCGGTGA
 321 ACTCCATCTCGTCCATGCCCTCGCCCGTGTACCAAGTGCAGGAAAGCCTTGCGCCTGAACATGGCAGTGAAGTGTCCGAG
 401 ATGCGCTTGAACAGCTCCTGGATGGCGGTGCTGTTGCCAATGAAGGTGGCTGACATCTTGAGGCCACGAGGAGGATGTC
 481 ACACACGGCCGTCTTACATTGTTGGGGATCCACTCCACGAAGTANCTTCTGTTCTTGTTCGCACGTTGAGCATCTGCT
 561 CATCCACCTCC (SEQ ID NO:16)

ACETA45

ACETA45 is a novel 619 bp gene fragment. The nucleic acid was initially identified in a 408 bp cloned fragment having the following sequence:

1 GGATCCTATGATCCTGAACGGCAGCCTGTGCTCTCTGTCTACCAGCCAGAGGACAACCTTGGAGGCTCTCCGAGACTCC
 81 CTGTAATCACCCTCTAGTGGGAGAGGTGGCAGAGGACAGGAGGAGGCTGAGAGGTTGCTCTGATGCTCATAGGTGT
 161 ATCATCAGTGCCTACGTGTCTATCCGGCTACTACCAAGACCCCTTCCAGGGCAAGATATCAGTCTAGCATCTCCTACAGA
 241 GGTATCCAGTCTCTAAGGAGGGGACAGTTTAGAAGACAGGGCAGTGACCCCGAAAACGTGCACAGCCAGGAGTAAGAAT
 321 CTCCATACCAGGTTGGGGATTTAGCTCAGTGGGTAGAGCGCTTGCCCTAAGGAAGTGAAGGCCCTGGGTTCGGTCCCCA
 401 GCTCCGGA (SEQ ID NO:17)

The cloned sequence was assembled into a contig resulting in the following consensus sequence:

1 TTTTTTTTTTTTTTACATAGGAAATAGTCTTTATTTGGTCTTCTGAAACGACAAACCAGAAATATAATTTGCCTTT
 81 AAAAACTTCTGTCTCAGGCTAAAGATATCACCAGTACAACACCCCTCCCTCCCCAGGCCCTTAGAAAAATCCCGGTTCT
 161 GGGATTGGCCACCAGCCAAAAGAGGAAGGAAGGCTGTGGCCAGCCTAGAGGATCCTATGATCCTGAACGGCAGCCTGT
 241 GCTCTCTGTCTACCAGCCAGAGGACAACCTTGGAGGCTCTCCGAGACTCCCTGTACTACCCCTGTAGTGGAGAAGGT
 321 GGCAGAGGACCAGGAGGAGGCTGAGAGGTTGCTCTGATGCTCATAGGTGTATCATCAGTGCCTACGTGTCTATCCGGCT
 401 ACTACCAAGACCCCTTCCAGGGCAAGATATCAGTCTAGCATCTCCTACAGAGGTATCCAGTCTCTAAGGAGGGGACAGTT
 481 TAGAAGACAGGGCAGTGACCCCGAAAACGTGCACAGCCAGGAGTAAGAATCTCCATACCAGGTTGGGGATTTAGCTCA
 561 GTGGGTAGAGCGCTTGCCCTAAGGAAGTGAAGGCCCTGGGTTCGGTCCCCAGCTCCGGA (SEQ ID NO:18)

ACETA46

ACETA46 is a novel 492 bp EST. The nucleic acid has the following sequence:

1 GGCATAGGATAGCTTCGTCTTAGCTTTCTACAATACTGTGTTTTAAGTCTACCCAGTAAACTAAGAGGCAGTACAACCTG
 81 AATACAACAGGAGGTGAGAACTCGGCAGGCCCTGGTGCTCAACAGTTCCGGAATAAACAGGTAGCGCGACCGG
 161 CATTAAGGTAGTTTGCAAATATTCAAGTTACACCGAGATTCCTTCAAACCTCTGCCTTGCTCAAAACAGCGCACCTG

241 GCCACTGCTAACACCGCGGCACGGCATTGGTGAATGGTGATGGTTTACTTTCTTCTTAATCAAAAGCCTGTGGAGGTG
321 CTGCCACGCAGCAGAGTAAGTCAGCTCTGCTCGAGGACGAGGTGCCAATTGTGAAGTCACAAAAGGTACAGTTTACATGA
401 GAAACGCACATTAAACAAGTGGCCCAAATACAGAGCTGGACAGAACAGACCAAGTANGTCTATAGTTACTCAACTGAACTA
481 ACATAAGTACAT (SEQ ID NO:19)

5

ACETA47

ACETA47 is a novel 413 bp EST. The nucleic acid has the following sequence:

1 GAAGGAGTGAATGTCCACTGGAGTTTATTTACAGACAACCTTAGGTAAGGCATTTTCCTCTAGGATCTACATCTTGTGAA
81 GTTACTTGGCTTCAGGCTTCTTGTCTCCAGCTTCAAGCTTGAGATGCTCAGGGGGCTGACGATAGGCAGGAAAGCCTGC
10 161 CAGGGGCTGTTCAAGTCAAACCTTGCGGAACTCTTGTGCCAACTCCACTGGCTAAGCCACTACCGGCTTAACCTGATCGGG
241 ATAACGTAGCTCAACATAGCCAGTGAGGGGAAAGTCTTCCGAAAGGATGTCCCTCGAAGCCATAATCTGTCAGGATCC
321 TTCTCAAGTCAGGGTGGTTGAAGAAGAAAACCCAACATGTCCAGACCTCCCTCTCATACCAATTGGCCGTTCACAACA
401 TCAGCCTCGTGCC (SEQ ID NO:20)

15

ACETA48

ACETA48 is a novel 138 bp gene fragment. The nucleic acid has the following sequence:

1 GGGCCCTGAGTTAGGTAACCATCTTGGCAAAGAACAGATACTCTCTCCATCCAAGTATTTCACTGTGGTTTCAATCAC
20 81 TTGGGCATTGGAATCTGTGGGGAGGAGGGCAGAATCTTTCTCATTCTTGTACA (SEQ ID NO:21)

ACETA49

ACETA49 is a novel 419 bp gene fragment. The nucleic acid has the following sequence:

25 1 AGATCTTTTCCAAATCCTCTGGCCATTCTGATGCTTGGTCTAGAGAATGTGGACCTTGCAGGAGAGCCCTGAGAATCT
81 CTCCCCAGAATAGAGCTGCCAGTCTGGCCAGCACTTAGCTTCTCATTTGGCTCACCCCTTGCTTTACAGATGTTTGTGTTG
161 TTTGTTGTTGTTTGTAGCTGTTGTGTTAACATTTCTACAGAATGCTTCTTGATAGGAGAGTAACAGAGGGCTGCCTGG
241 GTCTTAGGACAATCCATCTTTACCAGATGCCCTTGCCAAGCTCCGAAAGTCTCTTCAGGTAGAGCTAGGCTTCGCCAT
321 CAAGGCCACCGTATAAGCTGTGGCTTGAGGCTGCTGCTATCAGGAGGACAGGAGAAACAGGGTGGGGTGGAGTGTCTCAG
30 401 GAGAGCCATTGTGGGTACC (SEQ ID NO:22)

ACETA50

ACETA50 is a novel 109 bp gene fragment. The nucleic acid has the following

sequence:

1 CAATTGCAACCATAATGAAGATGCAGGAGTGAATGCCGCTGACCTGGGAGCCTGAGAAGTCATCCGTGTGTCCAGGT
81 GTCTTTGGGCACCACCCACATGGAGATCT (SEQ ID NO:23)

ACETA51

ACETA51 is a novel 299 bp gene fragment. The nucleic acid has the following

sequence:

1 AGATCTGGTCACCAACAATGAAGCATTGGCCACCCTTGTTCTGGGCCAGAAGAGTTCAAATGGCTTCAGGTGTCTCGGA
81 AGCTCCTTCCTATATTGGCCCTTGTCCTCCTTACAGATATGGAGATAGTGCCATGCAATGCGCCTGAACACGTCTTCCAG
161 TCCGTCGTTACCATGTCCACCAAGTGTGCTCTTGCTGGTCTTTGCCGTAGAGCCACAGGAAGCCTTGAATGTGCCTTG
241 CTCCCAAACATCCAAGGTCACCACCTCCTCCTCCAACCTCTGGCCCTGGTCGGCTAGC (SEQ ID NO:24)

ACETA52

ACETA52 is a novel 86 bp EST. The nucleic acid has the following sequence:

1 TGTTTGAAGATCGATCTCACTGGCATGGGGAACATATCTTGCTGCTCCCACTGGGCCTGTCTTTGCTCATGAGCTCCCT
81 GCTAGC (SEQ ID NO:25)

ACETA53

ACETA53 is a novel 111 bp EST. The nucleic acid has the following sequence:

1 GGGCCCTGAACATCATCAATGAAGAGAGAACAGCACTTCAAAGGAGGCCGTTGCAACAGTCCTGTCTCCCTGACCCTGAG
81 GAAGGACAACTTTTATATGCAAATATGTACA (SEQ ID NO:26)

ACETA54

ACETA54 is a novel 205 bp gene fragment. The nucleic acid has the following

sequence:

1 GGGCCCGGGGATGGGATGGACCCCAAGCTCCTAGGCTCCGCGCCTGGCTCAGAGGCTAACTGGCTTCGTAGGACGCAG
81 CTGACATCGCTGCCAGATGGCCTCCAAGCTGACCCCACTGACCCCTCCTGCTGCTGCTAGCTGGGGATAGACCTTC
161 TCAGATTCTGAAGTGACCAGCCACAGCTCCCAGGATCCACTAGT (SEQ ID NO:27)

GENERAL METHODS

The ACETA nucleic acids and encoded polypeptides can be identified using the information provided above. In some embodiments, the ACETA nucleic acids and polypeptide correspond to nucleic acids or polypeptides which include the various sequences
5 (referenced by SEQ ID NOs) disclosed for each ACETA polypeptide.

In its various aspects and embodiments, the invention includes providing a test cell population which includes at least one cell that is capable of expressing one or more of the sequences ACETA 1-170. By "capable of expressing" is meant that the gene is present in an intact form in the cell and can be expressed. Expression of one, some, or all of the ACETA
10 sequences is then detected, if present, and, preferably, measured. Using sequence information provided by the database entries for the known sequences, or the sequence information for the newly described sequences, expression of the ACETA sequences can be detected (if present) and measured using techniques well known to one of ordinary skill in the art. For example, sequences within the sequence database entries corresponding to ACETA sequences, or within
15 the sequences disclosed herein, can be used to construct probes for detecting ACETA RNA sequences in, *e.g.*, northern blot hybridization analyses or methods which specifically, and, preferably, quantitatively amplify specific nucleic acid sequences. As another example, the sequences can be used to construct primers for specifically amplifying the ACETA sequences in, *e.g.*, amplification-based detection methods such as reverse-transcription based polymerase
20 chain reaction. When alterations in gene expression are associated with gene amplification or deletion, sequence comparisons in test and reference populations can be made by comparing relative amounts of the examined DNA sequences in the test and reference cell populations.

Expression can be also measured at the protein level, *i.e.*, by measuring the levels of polypeptides encoded by the gene products described herein. Such methods are well known in
25 the art and include, *e.g.*, immunoassays based on antibodies to proteins encoded by the genes.

Expression level of one or more of the ACETA sequences in the test cell population is then compared to expression levels of the sequences in one or more cells from a reference cell population. Expression of sequences in test and control populations of cells can be compared using any art-recognized method for comparing expression of nucleic acid sequences. For

example, expression can be compared using GENE CALLING® methods as described in US Patent No. 5,871,697 and in Shimkets et al., Nat. Biotechnol. 17:798-803.

In various embodiments, the expression of one or more sequences encoding genes of related function, as listed in Table 1, is compared. These functions include, *e.g.*, “Protein
5 Production” (such as, ACETA 11-14), “Carbohydrate Metabolism” (ACETA 25-26), “Steroid Metabolism” (ACETA 22 and 24), and “Detoxification” (ACETA 33-34). In some embodiments, expression of members of two or more functional families are compared.

In various other embodiments, the expression of 2, 3, 4, 5, 6, 7, 8, 9, 10, 25, 50, 100, 150 or all of the sequences represented by ACETA 1-170 are measured. If desired, expression
10 of these sequences can be measured along with other sequences whose expression is known to be altered according to one of the herein described parameters or conditions.

The reference cell population includes one or more cells for which the compared parameter is known. The compared parameter can be, *e.g.*, hepatotoxic agent expression status. By “hepatotoxic agent expression status” is meant that it is known whether the
15 reference cell has had contact with a hepatotoxic agent. An example of a hepatotoxic agent is, *e.g.*, a nonsteroidal anti-inflammatory drug such as acetaminophen. Whether or not comparison of the gene expression profile in the test cell population to the reference cell population reveals the presence, or degree, of the measured parameter depends on the composition of the reference cell population. For example, if the reference cell population is
20 composed of cells that have not been treated with a known hepatotoxic agent, a similar gene expression level in the test cell population and a reference cell population indicates the test agent is not a hepatotoxic agent. Conversely, if the reference cell population is made up of cells that have been treated with a hepatotoxic agent, a similar gene expression profile between the test cell population and the reference cell population indicates the test agent is a
25 hepatotoxic agent.

In various embodiments, a ACETA sequence in a test cell population is considered comparable in expression level to the expression level of the ACETA sequence if its expression level varies within a factor of 2.0, 1.5, or 1.0 fold to the level of the ACETA transcript in the reference cell population. In various embodiments, a ACETA sequence in a
30 test cell population can be considered altered in levels of expression if its expression level

varies from the reference cell population by more than 1.0, 1.5, 2.0 or more fold from the expression level of the corresponding ACETA sequence in the reference cell population.

If desired, comparison of differentially expressed sequences between a test cell population and a reference cell population can be done with respect to a control nucleic acid whose expression is independent of the parameter or condition being measured. Expression levels of the control nucleic acid in the test and reference nucleic acid can be used to normalize signal levels in the compared populations.

In some embodiments, the test cell population is compared to multiple reference cell populations. Each of the multiple reference populations may differ in the known parameter. Thus, a test cell population may be compared to a first reference cell population known to have been exposed to a hepatotoxic agent, as well as a second reference population known have not been exposed to a hepatotoxic agent.

The test cell population that is exposed to, *i.e.*, contacted with, the test hepatotoxic agent can be any number of cells, *i.e.*, one or more cells, and can be provided *in vitro*, *in vivo*, or *ex vivo*.

In other embodiments, the test cell population can be divided into two or more subpopulations. The subpopulations can be created by dividing the first population of cells to create as identical a subpopulation as possible. This will be suitable, in, for example, *in vitro* or *ex vivo* screening methods. In some embodiments, various sub populations can be exposed to a control agent, and/or a test agent, multiple test agents, or, *e.g.*, varying dosages of one or multiple test agents administered together, or in various combinations.

Preferably, cells in the reference cell population are derived from a tissue type as similar as possible to test cell, *e.g.*, liver tissue. In some embodiments, the control cell is derived from the same subject as the test cell, *e.g.*, from a region proximal to the region of origin of the test cell. In other embodiments, the reference cell population is derived from a plurality of cells. For example, the reference cell population can be a database of expression patterns from previously tested cells for which one of the herein-described parameters or conditions (hepatotoxic agent expression status) is known.

The test agent can be a compound not previously described or can be a previously known compound but which is not known to be a hepatotoxic agent.

The subject is preferably a mammal. The mammal can be, *e.g.*, a human, non-human primate, mouse, rat, dog, cat, horse, or cow.

5 SCREENING FOR TOXIC AGENTS

In one aspect, the invention provides a method of identifying toxic agents, *e.g.*, hepatotoxic agents. The hepatotoxic agent can be identified by providing a cell population that includes cells capable of expressing one or more nucleic acid sequences homologous to those listed in Table 1. And Table 2 as ACETA 1-170. The sequences need not be identical to
10 sequences including ACETA 1-170, as long as the sequence is sufficiently similar that specific hybridization can be detected. Preferably, the cell includes sequences that are identical, or nearly identical to those identifying the ACETA nucleic acids shown in Table 1 and Table 2.

Expression of the nucleic acid sequences in the test cell population is then compared to the expression of the nucleic acid sequences in a reference cell population, which is a cell
15 population that has not been exposed to the test agent, or, in some embodiments, a cell population exposed the test agent. Comparison can be performed on test and reference samples measured concurrently or at temporally distinct times. An example of the latter is the use of compiled expression information, *e.g.*, a sequence database, which assembles information about expression levels of known sequences following administration of various
20 agents. For example, alteration of expression levels following administration of test agent can be compared to the expression changes observed in the nucleic acid sequences following administration of a control agent, such as acetaminophen.

An alteration in expression of the nucleic acid sequence in the test cell population compared to the expression of the nucleic acid sequence in the reference cell population that
25 has not been exposed to the test agent indicates the test agent is a hepatotoxic agent.

The invention also includes a hepatotoxic agent identified according to this screening method.

ASSESSING TOXICITY OF AN AGENT IN A SUBJECT

The differentially expressed ACETA sequences identified herein also allow for the hepatotoxicity of a toxic agent to be determined or monitored. In this method, a test cell population from a subject is exposed to a test agent, *i.e.* a hepatotoxic agent. If desired, test cell populations can be taken from the subject at various time points before, during, or after exposure to the test agent. Expression of one or more of the ACETA sequences, *e.g.*, ACETA: 1-170, in the cell population is then measured and compared to a reference cell population which includes cells whose hepatotoxic agent expression status is known. Preferably, the reference cells not been exposed to the test agent.

10 If the reference cell population contains no cells exposed to the treatment, a similarity in expression between ACETA sequences in the test cell population and the reference cell population indicates that the treatment is non-hepatotoxic. However, a difference in expression between ACETA sequences in the test population and this reference cell population indicates the treatment is hepatotoxic.

15 By "hepatotoxicity" is meant that the agent is damaging or destructive to liver when administered to a subject. In some embodiments, hepatotoxicity includes pericentral hepatic necrosis.

METHODS OF DIAGNOSING HEPATOTOXICITY

20 The invention further provides a method of diagnosing hepatotoxicity, in a subject. In this method, hepatotoxicity is diagnosed by examining the expression of one or more ACETA nucleic acid sequences from a test population of cells from a subject suspected to have been exposed to a hepatotoxic agent, *e.g.* non-steroidal anti-inflammatory drugs.

25 Expression of one or more of the ACETA nucleic acid sequences, *e.g.* ACETA: 1-170, or any combination of these sequences, is measured in the test cell and compared to the expression of the sequences in the reference cell population. The reference cell population contains at least one cell whose hepatotoxicity status is known. If the reference cell population contains cells that have not been exposed to a hepatotoxic agent, than a similarity in expression between ACETA sequences in the test population and the reference cell population indicates the subject does not have hepatotoxicity. A difference in expression between

ACETA sequences in the test population and the reference cell population indicates the reference cell population has hepatotoxicity.

Conversely, when the reference cell population contains cells that have been exposed to a hepatotoxic agent, a similarity in expression pattern between the test cell population and the reference cell population indicates the test cell population has hepatotoxicity. A difference in expression between ACETA sequences in the test population and the reference cell population indicates the subject does not have hepatotoxicity.

ACETA NUCLEIC ACIDS

Also provided in the invention are novel nucleic acid comprising a nucleic acid sequence selected from the group consisting of ACETA:1-10 and 43-54, or its complement, as well as vectors and cells including these nucleic acids.

Thus, one aspect of the invention pertains to isolated ACETA nucleic acid molecules that encode ACETA proteins or biologically active portions thereof. Also included are nucleic acid fragments sufficient for use as hybridization probes to identify ACETA-encoding nucleic acids (*e.g.*, ACETA mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of ACETA nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt) or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated ACETA nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of any of ACETA:1-10 and 43-54, or a complement of any of these nucleotide sequences, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of these nucleic acid sequences as a hybridization probe, ACETA nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to ACETA nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a

PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue.

Oligonucleotides comprise portions of a nucleic acid sequence having at least about 10 nt and as many as 50 nt, preferably about 15 nt to 30 nt. They may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in ACETA: 1-10 and 43-54. In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of these sequences, or a portion of any of these nucleotide sequences. A nucleic acid molecule that is complementary to the nucleotide sequence shown in ACETA:1-10 and 43-54 is one that is sufficiently complementary to the nucleotide sequence shown, such that it can hydrogen bond with little or no mismatches to the nucleotide sequences shown, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of ACETA:1-10 and 43-54 *e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically active portion of ACETA. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from

any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound
5 but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or
10 analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 45%, 50%, 70%, 80%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the
15 alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin
20 Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or
25 variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of a ACETA polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention,
30 homologous nucleotide sequences include nucleotide sequences encoding for a ACETA polypeptide of species other than humans, including, but not limited to, mammals, and thus

can include, *e.g.*, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding a human ACETA protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in a ACETA polypeptide, as well as a polypeptide having a ACETA activity. A homologous amino acid sequence does not encode the amino acid sequence of a human ACETA polypeptide.

The nucleotide sequence determined from the cloning of human ACETA genes allows for the generation of probes and primers designed for use in identifying and/or cloning ACETA homologues in other cell types, *e.g.*, from other tissues, as well as ACETA homologues from other mammals. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of a nucleic acid comprising a ACETA sequence, or an anti-sense strand nucleotide sequence of a nucleic acid comprising a ACETA sequence, or of a naturally occurring mutant of these sequences.

Probes based on human ACETA nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a ACETA protein, such as by measuring a level of a ACETA-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting ACETA mRNA levels or determining whether a genomic ACETA gene has been mutated or deleted.

"A polypeptide having a biologically active portion of ACETA" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of ACETA" can be prepared by isolating a portion of ACETA:1-10 and 43-54, that encodes a polypeptide having a ACETA biological activity, expressing the encoded portion of ACETA

protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of ACETA. For example, a nucleic acid fragment encoding a biologically active portion of a ACETA polypeptide can optionally include an ATP-binding domain. In another embodiment, a nucleic acid fragment encoding a biologically active portion of ACETA

5 includes one or more regions.

ACETA VARIANTS

The invention further encompasses nucleic acid molecules that differ from the disclosed or referenced ACETA nucleotide sequences due to degeneracy of the genetic code. These nucleic acids thus encode the same ACETA protein as that encoded by nucleotide
10 sequence comprising a ACETA nucleic acid as shown in, *e.g.*, ACETA:1-10 and 43-54

In addition to the rat ACETA nucleotide sequence shown in ACETA:1-10 and 43-54, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of a ACETA polypeptide may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the ACETA gene may exist
15 among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a ACETA protein, preferably a mammalian ACETA protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the ACETA gene. Any and all such nucleotide variations and resulting amino acid
20 polymorphisms in ACETA that are the result of natural allelic variation and that do not alter the functional activity of ACETA are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding ACETA proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of ACETA:1-10 and 43-54, are intended to be within the scope of the invention. Nucleic acid molecules
25 corresponding to natural allelic variants and homologues of the ACETA DNAs of the invention can be isolated based on their homology to the human ACETA nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a soluble human ACETA DNA can be isolated based on its homology to human

membrane-bound ACETA. Likewise, a membrane-bound human ACETA DNA can be isolated based on its homology to soluble human ACETA.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of ACETA:1-10 and 43-54. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250 or 500 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding ACETA proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.

Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other.

- 5 A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of
- 10 ACETA:1-10 and 43-54 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

- In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of ACETA:1-10 and 43-54 or fragments,
- 15 analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.*
- 20 (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

- In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of ACETA:1-10 and 43-54 or fragments, analogs or
- 25 derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other
- 30 conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN

MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo et al., 1981, *Proc Natl Acad Sci USA* 78: 6789-6792.

CONSERVATIVE MUTATIONS

5 In addition to naturally-occurring allelic variants of the ACETA sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced into an ACETA nucleic acid or directly into an ACETA polypeptide sequence without altering the functional ability of the ACETA protein. In some embodiments, the nucleotide sequence of ACETA:1-10 and 43-54 will be altered, thereby leading to changes in
10 the amino acid sequence of the encoded ACETA protein. For example, nucleotide substitutions that result in amino acid substitutions at various "non-essential" amino acid residues can be made in the sequence of ACETA:1-10 and 43-54. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of ACETA without altering the biological activity, whereas an "essential" amino acid residue is required for
15 biological activity. For example, amino acid residues that are conserved among the ACETA proteins of the present invention, are predicted to be particularly unamenable to alteration.

 In addition, amino acid residues that are conserved among family members of the ACETA proteins of the present invention, are also predicted to be particularly unamenable to alteration. As such, these conserved domains are not likely to be amenable to mutation. Other
20 amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the ACETA proteins) may not be essential for activity and thus are likely to be amenable to alteration.

 Another aspect of the invention pertains to nucleic acid molecules encoding ACETA proteins that contain changes in amino acid residues that are not essential for activity. Such
25 ACETA proteins differ in amino acid sequence from the amino acid sequences of polypeptides encoded by nucleic acids containing ACETA:1-10 and 43-54, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous, more preferably 60%, and still more preferably at least about 70%, 80%, 90%,
30 95%, 98%, and most preferably at least about 99% homologous to the amino acid sequence of

the amino acid sequences of polypeptides encoded by nucleic acids comprising ACETA:1-10 and 43-54.

5 An isolated nucleic acid molecule encoding a ACETA protein homologous to can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of a nucleic acid comprising ACETA:1-10 and 43-54, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into a nucleic acid comprising ACETA:1-10 and 43-54 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted
10 non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine,
15 asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in ACETA is replaced with another amino acid residue from the same side chain
20 family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a ACETA coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for ACETA biological activity to identify mutants that retain activity. Following mutagenesis of the nucleic acid, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

25 In one embodiment, a mutant ACETA protein can be assayed for (1) the ability to form protein:protein interactions with other ACETA proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant ACETA protein and a ACETA ligand; (3) the ability of a mutant ACETA protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to
30 bind ATP; or (5) the ability to specifically bind a ACETA protein antibody.

In other specific embodiments, the nucleic acid is RNA or DNA.

ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of a ACETA sequence or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire ACETA coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a ACETA protein, or antisense nucleic acids complementary to a nucleic acid comprising a ACETA nucleic acid sequence are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding ACETA. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding ACETA. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding ACETA disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of ACETA mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of ACETA mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of ACETA mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions

using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a ACETA protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified

such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of
5 antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the
10 strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

RIBOZYMES AND PNA MOIETIES

15 In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave ACETA mRNA transcripts to thereby
20 inhibit translation of ACETA mRNA. A ribozyme having specificity for a ACETA-encoding nucleic acid can be designed based upon the nucleotide sequence of a ACETA DNA disclosed herein. For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a ACETA-encoding mRNA. See, e.g., Cech *et al.* U.S. Pat. No. 4,987,071;
25 and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, ACETA mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, ACETA gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a ACETA nucleic acid (e.g., the
30 ACETA promoter and/or enhancers) to form triple helical structures that prevent transcription

of the ACETA gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of ACETA can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of ACETA can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of ACETA can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of ACETA can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of ACETA can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and

Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17:

- 5 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

- In other embodiments, the oligonucleotide may include other appended groups such as
- 10 peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See,
- 15 *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

ACETA POLYPEPTIDES

- 20 One aspect of the invention pertains to isolated ACETA proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-ACETA antibodies. In one embodiment, native ACETA proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another
- 25 embodiment, ACETA proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a ACETA protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

- An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from
- 30 which the ACETA protein is derived, or substantially free from chemical precursors or other

chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of ACETA protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of ACETA protein having less than about 30% (by dry weight) of non-ACETA protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-ACETA protein, still more preferably less than about 10% of non-ACETA protein, and most preferably less than about 5% non-ACETA protein. When the ACETA protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of ACETA protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of ACETA protein having less than about 30% (by dry weight) of chemical precursors or non-ACETA chemicals, more preferably less than about 20% chemical precursors or non-ACETA chemicals, still more preferably less than about 10% chemical precursors or non-ACETA chemicals, and most preferably less than about 5% chemical precursors or non-ACETA chemicals.

Biologically active portions of a ACETA protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the ACETA protein, *e.g.*, the amino acid sequence encoded by a nucleic acid comprising ACETA 1-10 and 43-54 that include fewer amino acids than the full length ACETA proteins, and exhibit at least one activity of a ACETA protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the ACETA protein. A biologically active portion of a ACETA protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a ACETA protein of the present invention may contain at least one of the above-identified domains conserved between the ACETA proteins. An

alternative biologically active portion of a ACETA protein may contain at least two of the above-identified domains. Another biologically active portion of a ACETA protein may contain at least three of the above-identified domains. Yet another biologically active portion of a ACETA protein of the present invention may contain at least four of the above-identified domains.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native ACETA protein.

In some embodiments, the ACETA protein is substantially homologous to one of these ACETA proteins and retains its the functional activity, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below.

In specific embodiments, the invention includes an isolated polypeptide comprising an amino acid sequence that is 80% or more identical to the sequence of a polypeptide whose expression is modulated in a mammal to which hepatotoxic agent is administered.

15 DETERMINING HOMOLOGY BETWEEN TWO OR MORE SEQUENCES

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See *Needleman and Wunsch* 1970 *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty

of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of a DNA sequence comprising ACETA: 1-10 and 43-54..

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

CHIMERIC AND FUSION PROTEINS

The invention also provides ACETA chimeric or fusion proteins. As used herein, an ACETA "chimeric protein" or "fusion protein" comprises an ACETA polypeptide operatively linked to a non-ACETA polypeptide. A "ACETA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to ACETA, whereas a "non-ACETA polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the ACETA protein, *e.g.*, a protein that is different from the ACETA protein and that is derived from the same or a different organism. Within an ACETA fusion protein the ACETA polypeptide can correspond to all or a portion of an ACETA protein. In one embodiment, an ACETA fusion protein comprises at least one biologically active portion of an ACETA protein. In another embodiment, an ACETA fusion protein comprises at least two biologically active portions of an ACETA protein. In yet another embodiment, an ACETA fusion protein comprises at least three biologically active portions of an ACETA protein. Within the fusion protein, the term "operatively linked" is intended to

indicate that the ACETA polypeptide and the non-ACETA polypeptide are fused in-frame to each other. The non-ACETA polypeptide can be fused to the N-terminus or C-terminus of the ACETA polypeptide.

For example, in one embodiment an ACETA fusion protein comprises an ACETA
5 domain operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds which modulate ACETA activity (such assays are described in detail below).

In yet another embodiment, the fusion protein is a GST-ACETA fusion protein in which the ACETA sequences are fused to the C-terminus of the GST (*i.e.*, glutathione
10 S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant ACETA.

In another embodiment, the fusion protein is an ACETA protein containing a heterologous signal sequence at its N-terminus. For example, a native ACETA signal sequence can be removed and replaced with a signal sequence from another protein. In certain
15 host cells (*e.g.*, mammalian host cells), expression and/or secretion of ACETA can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an ACETA-immunoglobulin fusion protein in which the ACETA sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The
20 ACETA-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ACETA ligand and a ACETA protein on the surface of a cell, to thereby suppress ACETA-mediated signal transduction *in vivo*. The ACETA-immunoglobulin fusion proteins can be used to affect the bioavailability of an ACETA cognate ligand. Inhibition of the
25 ACETA ligand/ACETA interaction may be useful therapeutically for both the treatments of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the ACETA-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-ACETA antibodies in a subject, to purify ACETA ligands, and in screening assays to identify molecules that inhibit the interaction of ACETA
30 with a ACETA ligand.

An ACETA chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An ACETA-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the ACETA protein.

ACETA AGONISTS AND ANTAGONISTS

The present invention also pertains to variants of the ACETA proteins that function as either ACETA agonists (mimetics) or as ACETA antagonists. Variants of the ACETA protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the ACETA protein. An agonist of the ACETA protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the ACETA protein. An antagonist of the ACETA protein can inhibit one or more of the activities of the naturally occurring form of the ACETA protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the ACETA protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the ACETA proteins.

Variants of the ACETA protein that function as either ACETA agonists (mimetics) or as ACETA antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*,

truncation mutants, of the ACETA protein for ACETA protein agonist or antagonist activity. In one embodiment, a variegated library of ACETA variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of ACETA variants can be produced by, for example, enzymatically ligating a mixture
5 of synthetic oligonucleotides into gene sequences such that a degenerate set of potential ACETA sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of ACETA sequences therein. There are a variety of methods which can be used to produce libraries of potential ACETA variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene
10 sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential ACETA sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura
15 *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

POLYPEPTIDE LIBRARIES

In addition, libraries of fragments of the ACETA protein coding sequence can be used to generate a variegated population of ACETA fragments for screening and subsequent selection of variants of an ACETA protein. In one embodiment, a library of coding sequence
20 fragments can be generated by treating a double stranded PCR fragment of a ACETA coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the
25 resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the ACETA protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene
30 products having a selected property. Such techniques are adaptable for rapid screening of the

gene libraries generated by the combinatorial mutagenesis of ACETA proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the

5 combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify ACETA variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering

10 6:327-331).

ANTI-ACETA ANTIBODIES

An isolated ACETA protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind ACETA using standard techniques for polyclonal and monoclonal antibody preparation. The full-length ACETA protein can be used or,

15 alternatively, the invention provides antigenic peptide fragments of ACETA for use as immunogens. The antigenic peptide of ACETA comprises at least 8 amino acid residues of the amino acid sequence encoded by a nucleic acid comprising the nucleic acid sequence shown in ACETA:1-10 and 43-54 and encompasses an epitope of ACETA such that an antibody raised against the peptide forms a specific immune complex with ACETA.

20 Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of ACETA that are located on the surface of the protein, *e.g.*, hydrophilic regions. As a means for targeting antibody production, hydropathy plots showing regions of

25 hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each incorporated herein by reference in their entirety.

30 ACETA polypeptides or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these

protein components. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and F_{(ab')₂} fragments, and an F_{ab} expression library. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to an ACETA protein sequence, or derivatives, fragments, analogs or homologs thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly expressed ACETA protein or a chemically synthesized ACETA polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against ACETA can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of ACETA. A monoclonal antibody composition thus typically displays a single binding affinity for a particular ACETA protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular ACETA protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (see Kohler & Milstein, 1975 *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (see Kozbor, *et al.*, 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, *et al.*, 1985 In:

MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, *et al.*, 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a ACETA protein (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, *et al.*, 1989 *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a ACETA protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. See *e.g.*, U.S. Patent No. 5,225,539. Antibody fragments that contain the idiotypes to a ACETA protein may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab)_2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab)_2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Additionally, recombinant anti-ACETA antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Pat. No. 4,816,567; European Patent Application No. 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS* 84:214-218; Nishimura *et al.* (1987) *Cancer Res* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; Shaw *et al.* (1988) *J Natl Cancer Inst.* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; U.S. Pat. No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525;

Verhoeven *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J Immunol* 141:4053-4060.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and
5 other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a ACETA protein is facilitated by generation of hybridomas that bind to the fragment of a ACETA protein possessing such a domain. Antibodies that are specific for one or more domains within a ACETA protein, *e.g.*, domains spanning the above-identified conserved regions of ACETA
10 family proteins, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-ACETA antibodies may be used in methods known within the art relating to the localization and/or quantitation of a ACETA protein (*e.g.*, for use in measuring levels of the ACETA protein within appropriate physiological samples, for use in diagnostic methods, for
15 use in imaging the protein, and the like). In a given embodiment, antibodies for ACETA proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds [hereinafter "Therapeutics"].

An anti-ACETA antibody (*e.g.*, monoclonal antibody) can be used to isolate ACETA
20 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-ACETA antibody can facilitate the purification of natural ACETA from cells and of recombinantly produced ACETA expressed in host cells. Moreover, an anti-ACETA antibody can be used to detect ACETA protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the ACETA protein. Anti-ACETA
25 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive
30 materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group

complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include
5 luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

ACETA RECOMBINANT EXPRESSION VECTORS AND HOST CELLS

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding ACETA protein, or derivatives, fragments, analogs or
10 homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a linear or circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of
15 autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are
20 operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses,
25 adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively linked to the nucleic acid
30 sequence to be expressed. Within a recombinant expression vector, "operably linked" is

intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, ACETA proteins, mutant forms of ACETA, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of ACETA in prokaryotic or eukaryotic cells. For example, ACETA can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification

of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the ACETA expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *EMBO J* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

Alternatively, ACETA can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith *et al.* (1983) *Mol Cell Biol* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J*

6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells. See, *e.g.*, Chapters 16 and 17 of Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to ACETA mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant

plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," Reviews--Trends in Genetics, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, ACETA protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer

resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding ACETA or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable
5 marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an ACETA protein. Accordingly, the invention further provides methods for producing ACETA protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a
10 recombinant expression vector encoding ACETA has been introduced) in a suitable medium such that ACETA protein is produced. In another embodiment, the method further comprises isolating ACETA from the medium or the host cell.

PHARMACEUTICAL COMPOSITIONS

The ACETA nucleic acid molecules, ACETA proteins, and anti-ACETA antibodies
15 (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings,
20 antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5%
25 human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, 5 intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates, and agents 10 for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous 15 solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under 20 the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the 25 maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged 30 absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a ACETA protein or anti-ACETA antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays

or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal
5 delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides,
10 polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared
15 according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each
20 unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment
25 of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Pat. No. 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of
30 the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively,

where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

5 The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

KITS AND NUCLEIC ACID COLLECTIONS FOR IDENTIFYING ACETA NUCLEIC ACIDS

In another aspect, the invention provides a kit useful for examining hepatotoxicity of agents. The kit can include nucleic acids that detect two or more ACETA sequences. In preferred embodiments, the kit includes reagents which detect 3, 4, 5, 6, 8, 10, 12, 15, 20, 25,
10 50, 100 or all of the ACETA nucleic acid sequences.

The invention also includes an isolated plurality of sequences which can identify one or more ACETA responsive nucleic acid sequences.

The kit or plurality may include, *e.g.*, sequence homologous to ACETA nucleic acid sequences, or sequences which can specifically identify one or more ACETA nucleic acid
15 sequences.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other
20 aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

1. A method of screening a test agent for hepatotoxicity, the method comprising:
 - (a) providing a test cell population comprising a cell capable of expressing one or more nucleic acid sequences selected from the group consisting of ACETA: 1-169 and 170;
 - (b) contacting the test cell population with a test agent;
 - (c) measuring expression of one or more of the nucleic acid sequences in the test cell population;
 - (d) comparing the expression of the nucleic acid sequences in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell whose exposure status to a hepatotoxic agent is known; and
 - (e) identifying a difference in expression levels of the ACETA sequence, if present, in the test cell population and reference cell population,thereby screening said test agent for hepatotoxicity.
2. The method of claim 1, wherein the method comprises comparing the expression of 15 or more of the nucleic acid sequences.
3. The method of claim 1, wherein the expression of the nucleic acid sequences in the test cell population is decreased as compared to the reference cell population.
4. The method of claim 1, wherein the expression of the nucleic acid sequences in the test cell population is increased as compared to the reference cell population.

- 5 The method of claim 1, wherein the test cell population is provided *in vitro*.
- 6 The method of claim 1, wherein the test cell population is provided *ex vivo* from a mammalian subject.
- 7 The method of claim 1, wherein the test cell population is provided *in vivo* in a mammalian subject.
- 8 The method of claim 1, wherein the test cell population is derived from a human or rodent subject.
- 9 The method of claim 1, wherein the test cell population includes a hepatocyte.
10. The method of claim 1, wherein said test agent is a nonsteroidal antiinflammatory.
11. The method of claim 10, wherein the nonsteroidal antiinflammatory drug is acetaminophen.
12. A method of assessing the hepatotoxicity of a test agent in a subject, the method comprising:
 - (a) providing from the subject a test cell population comprising a cell capable of expressing one or more nucleic acid sequences selected from the group consisting of ACETA: 1-169 and 170;
 - (b) contacting the test cell population with a test agent;
 - (c) measuring expression of one or more of the nucleic acid sequences in the test cell population; and

(d) comparing the expression of the nucleic acid sequences in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell whose exposure status to a hepatotoxic agent is known;

(e) identifying a difference in expression levels of the nucleic acid sequences, if present, in the test cell population and the reference cell population,

thereby assessing the hepatotoxicity of the test agent in the subject.

13. The method of claim 12, wherein the expression of the nucleic acid sequences in the test cell population is decreased as compared to the reference cell population.
14. The method of claim 12, wherein the expression of the nucleic acid sequences in the test cell population is increased as compared to the reference cell population.
15. The method of claim 12, wherein said subject is a human or rodent.
16. The method of claim 12, wherein the test cell population is provided *ex vivo* from said subject.
17. The method of claim 12, wherein the test cell population is provided *in vivo* from said subject.
18. A method of diagnosing or determining the susceptibility to hepatotoxicity in a subject, the method comprising:
 - (a) providing from the subject a test cell population comprising cells capable of expressing on or more nucleic acid sequences selected from the group consisting of HEPATO: 1-169, and 170;

- (b) measuring expression of one or more of the nucleic acid sequences in the test cell population; and
 - (c) comparing the expression of the nucleic acid sequences in the test cell population to the expression of the nucleic acid sequences in a reference cell population comprising at least one cell from a subject not suffering from hepatotoxicity; and
 - (d) identifying a difference in expression levels of the nucleic acid sequences, if present, in the test cell population and reference cell population, thereby diagnosing or determining the susceptibility hepatotoxicity in the subject.
19. The method of claim 18, wherein said hepatotoxicity results in pericentral hepatic necrosis in said subject.
20. An isolated nucleic acid comprising a nucleic acid sequence selected from the group consisting of an ACETA 1-10, 43-53, and 54 nucleic acid, or its complement.
21. A vector comprising the nucleic acid of claim 20.
22. A cell comprising the vector of claim 21.
23. A polypeptide encoded by the nucleic acid of claim 20.
24. An antibody which specifically binds to the polypeptide of claim 23.
25. A kit which detects two or more of the nucleic acid sequences selected from the group consisting of ACETA: 1-169 and 170.

26. An array which detects one or more of the nucleic acid selected from the group consisting of ACETA: 1-169 and 170.

27. A plurality of nucleic acid comprising one or more of the nucleic acid selected from the group consisting of ACETA: 1-169 and 170.